

METHOD AND DEVICE FOR EXTENDING CHROMOSOME ON SLIDE GLASS

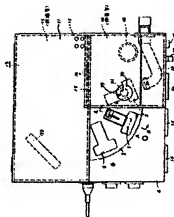
Publication number: JP3209163
 Publication date: 1991-09-12
 Inventor: MIYASAKA CHIAKI; TATEYA HIROE
 Applicant: CHIYODA SEISAKUSHO; SAKURA FINETECHNICAL CO LTD
 Classification:
 - International: G01N33/48; G01N1/28; G01N33/48; G01N1/28; (IPC1-7): G01N33/48
 - European:
 Application number: JP19900002422 19900111
 Priority number(s): JP19900002422 19900111

Report a data error here

Abstract of JP3209163

PURPOSE: To perform automatic treatment by providing moving and humidifying means, an extending chamber, and a drying chamber to perform the stage of chromosome at a high temperature and the stage of evaporation of a Camot's fixative and water at a low humidity after dropping of the fixative including a cell nucleus on the upper face of a wet slide glass.

CONSTITUTION: A slide glass 2 is rotated together with a turntable 3 and is put in an extending chamber 14 and is wetted with minute water drops sprayed from a hose 22, and a disk 30 is rotated to drop the Camot's fixative including a cell nucleus on the upper face from a circular hole 21, and then, the membrane is broken and the chromosome is extended together with the fixative. The slide glass 2 is put in a heated drying chamber at a low humidity in accordance with rotation of the table 3, and water drops and the fixative on the upper face are evaporated and only the chromosome remains. A cover glass is so stuck that the chromosome on the slide glass 2 is covered with it. Thus, the chromosome is easily extended without skill and the work is automatically performed.



Data supplied from the esp@cenet database - Worldwide

(51) Int.Cl.⁷

識別記号

F I

G 0 6 F 9/44

5 5 0

G 0 6 F 9/44

5 5 0 N

請求項の数 5 (全 19 頁)

(21) 出願番号 特願平9-273797

(22) 出願日 平成9年9月19日(1997.9.19)

(65) 公開番号 特開平11-96010

(43) 公開日 平成11年4月9日(1999.4.9)

審査請求日 平成9年9月19日(1997.9.19)

(73) 特許権者 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72) 発明者 大野 和彦

東京都港区芝五丁目7番1号 日本電気株式会社内

(74) 代理人 100080816

弁理士 加藤 朝道

審査官 久保 光宏

最終頁に続く

(54) 【発明の名称】 分類装置

(57) 【特許請求の範囲】

【請求項1】 分類手がかりとなる条件部と分類結果の組を事例データとして、該事例データのなかで分類結果が既知である既知事例データを用いて分類ルールを自動生成し、分類結果が未知である未知事例データを、前記分類ルールを用いて自動分類する分類装置において、既知事例データおよび未知事例データを入力する入力手段と、確率的な情報を含む分類ルールを蓄える分類ルールデータベースと、前記既知事例データを条件部の論理関係に基づいてネットワーク化して蓄積する事例データベースと、入力された前記既知事例データおよび前記未知事例データの条件部と前記分類ルールを用いて分類結果の確率値を推定する確率値推定手段と、

分類ルールの有効性を統計的検定により評価することにより無駄な分類ルールの生成を抑制する分類ルール生成手段と、を備え、

前記分類ルール生成手段が、前記入力手段から入力された前記既知事例データの分類結果の確率値と、前記確率値推定手段から出力される分類結果の確率値と、を比較して、分類ルールを生成するか抑制するかを判定する、ことを特徴とする分類装置。

【請求項2】 分類手がかりとなる条件部と分類結果の組を事例データとして、該事例データのなかで分類結果が既知である既知事例データを用いて分類ルールを自動生成し、分類結果が未知である未知事例データを、前記分類ルールを用いて自動分類する分類装置において、既知事例データおよび未知事例データを入力する入力手

段と、

確率的な情報を含む分類ルールを蓄える分類ルールデータベースと、

前記既知事例データを条件部の論理関係に基づいてネットワーク化して蓄積する事例データベースと、

入力された前記既知事例データおよび前記未知事例データの条件部と前記分類ルールを用いて分類結果の確率値を推定する確率値推定手段と、

分類ルールの有効性を統計的検定により評価することにより無駄な分類ルールの生成を抑制する分類ルール生成手段と、

入力された前記既知事例データの条件部全部あるいは条件の一部を入力して受取り、前記事例データベースを探索して否定条件を含む事例ノードを推定する否定条件探索手段を備え、

前記否定条件探索手段を用いて、否定条件を含む分類ルールを生成する機能を追加した分類ルール生成手段を備える、ことを特徴とする分類装置。

【請求項3】 分類手がかかりとなる条件部と分類結果の組を事例データとして、該事例データのなかで分類結果が既知である既知事例データを用いて分類ルールを自動生成し、分類結果が未知である未知事例データを、前記分類ルールを用いて自動分類する処理を、コンピュータで実行させるプログラムを記録した記録媒体において、前記コンピュータは、既知事例データおよび未知事例データを入力する入力手段と、

確率的な情報を含む分類ルールを蓄える分類ルールデータベースと、

前記既知事例データを条件部の論理関係に基づいてネットワーク化して蓄積する事例データベースと、を備え、

入力された前記既知事例データについて前記分類ルールデータベースから新しい分類ルールを生成する処理と、分類ルールの有効性を統計的検定により評価して無駄なルールの生成を抑制するルール有効性判定処理と、

入力された前記既知事例データの条件部あるいは条件の一部を用いて前記事例データベースを探索して否定条件を含む事例ノードを推定する否定条件の探索手段と、

該否定条件の探索手順で推定された否定条件を含む分類ルールを生成する否定条件付き分類ルールの生成処理と、

を前記コンピュータに実行させるためのプログラムを記録した記録媒体。

【請求項4】 分類手がかかりとなる条件部と分類結果の組を事例データとして、該事例データのなかで分類結果が既知である既知事例データを用いて分類ルールを自動生成し、分類結果が未知である未知事例データを前記分類ルールを用いて自動分類する処理を、コンピュータで実行させるプログラムを記録した記録媒体において、前記コンピュータは、既知事例データおよび未知事例データを入力する入力装

置と、

確率的な情報を含む分類ルールを蓄える分類ルールデータベースと、

前記既知事例データを条件部の論理関係に基づいてネットワーク化して蓄積する事例データベースと、を含み、

(a) 前記入力装置から入力された前記既知事例データおよび前記未知事例データの条件部と前記分類ルールを用いて分類結果の確率値を推定する確率値推定手段、及び、

(b) 分類ルールの有効性を統計的検定により評価することにより無駄な分類ルールの生成を抑制する分類ルール生成手段であって、前記入力装置から入力された前記既知事例データの分類結果の確率値と、前記確率値推定手段から出力される分類結果の確率値と、を比較して、分類ルールを生成するか抑制するかを判定する分類ルール生成手段、

の上記 (a)、(b) の各手段を前記コンピュータ上で機能させるためのプログラムを記録した記録媒体。

【請求項5】 分類手がかかりとなる条件部と分類結果の組を事例データとして、該事例データのなかで分類結果が既知である既知事例データを用いて分類ルールを自動生成し、分類結果が未知である未知事例データを前記分類ルールを用いて自動分類する処理を、コンピュータで実行させるプログラムを記録した記録媒体において、前記コンピュータは、

既知事例データおよび未知事例データを入力する入力装置と、

確率的な情報を含む分類ルールを蓄える分類ルールデータベースと、

前記既知事例データを条件部の論理関係に基づいてネットワーク化して蓄積する事例データベースと、を含み、

(a) 前記入力装置から入力された前記既知事例データおよび前記未知事例データの条件部と前記分類ルールを用いて分類結果の確率値を推定する手段、

(b) 入力された前記既知事例データの条件部全部あるいは条件の一部を入力して受取り、前記事例データベースを探索して否定条件を含む事例ノードを推定する否定条件探索手段、及び、

(c) 分類ルールの有効性を統計的検定により評価することにより無駄な分類ルールの生成を抑制する手段であって、前記否定条件探索手段を用いて、否定条件を含む分類ルールを生成する機能を備えた分類ルール生成手段、

の上記 (a) ~ (c) の各手段を前記コンピュータ上で機能させるためのプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、分類済の事例データを用いて分類ルールを自動生成し、未分類の事例データを自動分類する分類装置に関する。

【0002】

【従来の技術】分類事例から分類ルールを自動生成し、これを用いて分類を行う方法としては、統計的な手法、知識処理的な手法、計算論的学習理論に基づく手法、ニューラルネット的な手法など、多くの方法が提案されている。これらの手法は、各々長所、短所があり、適する問題領域も変わって来る。

【0003】統計的な手法は、確率モデルを基礎に事例データの解析を行う事を主目的にするため、一般的には事例データに潜在する主要な傾向を見つ出すために利用されることが多い。また、ルール生成に必要な計算量も少ない。半面、例外的な分類処理は得意でないことが多い。代表的手法は、数量化II類やベイズ判別などが知られている。

【0004】知識処理的な手法、および計算論的学習理論は、機械学習の実現を目指した研究の過程で提案されている手法および理論である。機械学習とは、簡単に言えば、計算機が自律的に適応的な知識（情報、ルール、プログラム等）を自動生成することである。自動分類もこの機械学習機能の一部として位置付けられている。統計的手法の主目的は事例データの主要な傾向の発見であるが、機械学習の場合は、最終目的が計算機の知能化にあるため、目的が多様化している。研究者（グループ）毎に目的は少しずつ異なっており、例えば、事例と矛盾しない仮説の範囲を求めること、例外的に振舞う事例データを認識しこれを処理すること、あるいは事例を生成するプログラムを自動生成することなどが目的になる。

【0005】ニューラルネットの手法は、擬似的な神経細胞を用いた模倣システムであり、学習やパターン識別に应用することができる。簡単に使える半面、分類ルールがブラックボックスとなり人間によるチェックや修正が難しいという問題点がある。

【0006】以下では、従来技術として、帰納理論装置（特公平07-43722号公報参照）を注目する。上記特公平07-43722号公報に記載の帰納理論装置は、諸条件の組合せによって成立する事象に対する一般的な知識を、この対象分野の事例から帰納的に求めることにより、知識処理に必要な知識情報を効果的に得ることを目的としたものである。

【0007】この帰納理論装置は、事例データの集合を入力されたのち、各分類結果を成立させる十分条件と必要条件を自動生成する。これを用いて、未知の条件が与えられた際、十分条件が満たされていれば当該分類結果は成立と判定し、必要条件が満たされていなければ当該分類結果は不成立と判定する。

【0008】

【発明が解決しようとする課題】この技術の問題点は、事例データの組が論理的に分類不能な場合に、有効な分類ルールを生成できないことである。例えば、以下の事

例データが与えられたとしよう。

【0009】

【表1】

| 事例 | 分類結果 | 条件 |
|-----|------|------|
| 鰻 | 魚 | 水中生活 |
| 蛙 | 魚 | 水中生活 |
| 鱒 | 魚 | 水中生活 |
| イルカ | 獣 | 水中生活 |

【0010】上記の事例データから分類ルールが生成された後、分類結果が未知の事例データ

【0011】

【表2】

| 事例 | 分類結果 | 条件 |
|----|------|------|
| 鰻 | ? | 水中生活 |

【0012】を分類する場合を考える。

【0013】条件“水中生活”は、分類結果“魚”を導く十分条件でないため、鰻は魚とは分類できない。また、“水中生活”は“魚”の必要条件であるため、鰻が魚であることは否定できない。つまり、上記の事例データの集合からは、“水中生活”と“魚”に関する積極的な分類ルールは生成されない。この事情は、“獣”に関しても同様である。

【0014】以上は、論理的には当然のことであるが、現実的な応用の面からは問題点がある。

【0015】第1の問題点は、現実的な応用問題では論理的に分類不能な事例データが与えられることが多いことである。

【0016】第2の問題点は、前記の帰納理論装置は事例データの情報を十分に活用しているとは言えない点である。例えば、水中生活をほとんどの事例が“魚”であり、それらが学習用の事例データとしてシステムに与えられた場合、“確実ではないが、水中生活をするのは魚であろう”という分類ルールができることが望ましい。この際、確実さの程度が定量的にかわることも重要である。

【0017】前記の帰納理論装置と対比させた場合、本発明が解決しようとする課題は、

- ・論理的に分類不能な事例データの入力に対して、確実さの程度を定量的に含む分類ルールを生成し、
- ・そのような分類ルールを用いて、未知の事例を確実さの定量的値で分類することである。

【0018】本発明の目的について、より一般的な立場から述べると、現実的な分類問題に対して、有効な学習型の自動分類を行う装置を提供することにある。

【0019】ここで、現実的な分類問題とは、前記のような論理的には分類不能な事例を含み、かつ、条件や分類結果の種類が大きな問題のことを意味する。そして、これを有効に分類するとは、

1. 現実的な時間と計算資源内で分類に有効な分類ルー

ルを生成でき、

2. 分類結果を確実さの定量化つきで出力できる、ことをいう。

【0020】

【課題を解決するための手段】前記目的を達成するため、本発明は、その概略を述べれば、確実さの定量化として確率値を分類ルールに導入し、事例を用いて確率値付きの分類ルールを生成し、この分類ルールを用いて確率的な分類を行うようにしたものである。

【0021】より詳細には、本発明は、分類手がかりとなる条件部と分類結果の組を事例データとして、該事例データのなかで分類結果が既知である既知事例データを用いて分類ルールを自動生成し、分類結果が未知である未知事例データを、前記分類ルールを用いて自動分類する分類装置において、既知事例データおよび未知事例データを入力する入力手段と、確率的な情報を含む分類ルールを蓄える分類ルールデータベースと、前記既知事例データを条件部の論理関係に基づいてネットワーク化して蓄積する事例データベースと、入力された前記既知事例データおよび前記未知事例データの条件部と前記分類ルールを用いて分類結果の確率値を推定する確率値推定手段と、分類ルールの有効性を統計的検定により評価することにより無駄な分類ルールの生成を抑制する分類ルール生成手段と、を備えたことを特徴とする。

【0022】また、本発明は、前記入力手段から入力された前記既知事例データの分類結果の確率値と、前記確率値推定手段から出力される分類結果の確率値と、を比較して、その差が予め定めた値よりも大きな場合に、分類ルールを生成する、ことを特徴とする。

【0023】また、本発明は、入力された前記既知事例データの条件部全部あるいは条件の一部を入力として受取り、前記事例データベースを探索して否定条件を含む事例ノードを推定する否定条件探索手段を備え、前記否定条件探索手段を用いて、否定条件を含む分類ルールを

生成する機能を追加した分類ルール生成手段を備える、ことを特徴とする。

【0024】本発明は、第1の手段として、分類ルールの生成時に、その分類ルールの有効性を判定し、無駄な分類ルールの生成を防ぐ（請求項1参照）。

【0025】本発明は、第2の手段として、条件の論理的関係に基づいてネットワーク化された事例データベースを探索することにより否定条件を発見し、否定条件付きの分類ルールを生成する（請求項3参照）。

【0026】本発明によれば、上記第1の手段により、分類ルールの記憶のために必要な記憶装置の総量を減少させることができ、分類ルールの利用・操作の処理時間を短縮させることができる。これは、現実問題を解く際に有効である。

【0027】また上記第2の手段により、全事例を探索するよりも短い処理時間で否定条件の探索を行うことができ、否定条件付きの分類ルールを効率的に生成することができる。否定条件付きの分類ルールは自動分類の精度向上につながる。

【0028】【発明の概要】本発明の実施の形態を説明する前に、本発明の原理について説明する。

【0029】本発明の特徴は、事例データから分類ルールを自動生成する点にある。この処理の背景にある確率統計的な考え方を下段に事例データ、母集団、分類ルール等について説明する。

【0030】事例データ：まず、事例データを定義する。

【0031】事例データとは、条件部と分類結果の対である。条件部は、一般に、複数の条件の論理積から構成される。これらの条件は、該事例データに関して成立している条件である。また、成立している条件は条件部に含まれないものとする。以下は条件部の例である。

【0032】

【数1】

… (1)

【0038】

【表3】

| 事例名称 | 分類結果 | 条件部 |
|------|------|-------------------|
| 鰻1 | 魚 | 水中生活 ∧ えら呼吸 ∧ 鱗あり |

【0039】この事例データは、水中生活を営み、えら呼吸を行い、鱗があり、魚に分類され、鰻1と呼ばれる。

【0040】<事例データ集合>：ある時点までに投入された事例データの集まりを事例データ集合と呼ぶ。以下は、事例データ集合の例である。

【0041】

【表4】

水中生活 ∧ えら呼吸 ∧ 鱗あり

【0033】但し、記号

【0034】

【数2】

【0035】は論理積を表す記号である。上記の例では、「水中生活」、「えら呼吸」、「鱗あり」の全ての条件が満たされている事を示す。

【0036】分類結果は、離散的な記号である。以下は分類結果の例である。

魚

【0037】条件部と分類結果の対により事例データが構成される。また、本質的ではないが、説明のため事例名称を追加する場合もある。以下に1件の事例データの例を示す。

| 事例名称 | 分類結果 | 条件部 |
|------|------|--------------------|
| 鱈3 | 魚 | 水中生活 A えら呼吸 |
| 鮭1 | 魚 | 水中生活 A えら呼吸 A 潮上産卵 |
| 鰐1 | 獣 | 陸上生活 A 肺呼吸 |
| 鰻2 | 魚 | 水中生活 A えら呼吸 A 小型 |

【0042】<事例データの観測確率>: 事例データにはそれが観測される確率が付属していると考えられる (これは

は仮定である)。この観測確率は、一般に0以上1以下の実数値を取り、全事例データの観測確率の和は1となる。観測確率は「測度」と呼ばれることもある。

【0043】観測確率も含めた事例データ集合を母集団と呼ぶ。母集団の例を示す。

【0044】

【表5】

| 事例名称 | 観測確率 | 分類結果 | 条件部 |
|------|------|------|--------------------|
| 鱈3 | 0.3 | 魚 | 水中生活 A えら呼吸 |
| 鮭1 | 0.2 | 魚 | 水中生活 A えら呼吸 A 潮上産卵 |
| 鰐1 | 0.1 | 獣 | 陸上生活 A 肺呼吸 |
| 鰻2 | 0.4 | 魚 | 水中生活 A えら呼吸 A 小型 |

【0045】事例データは母集団の観測確率に基づいて発生すると仮定する。つまり、表1の母集団から事例データを1件観測した場合、それが

【0046】

【表6】

| 事例名称 | 分類結果 | 条件部 |
|------|------|------------|
| 鰐1 | 獣 | 陸上生活 A 肺呼吸 |

【0047】である確率は0.1である。

【0048】<分類結果の条件つき確率>: 事例データにおいて、条件Cが成立するとは、

1. 事例データの条件部に、条件Cの肯定条件が全て含まれ、

2. 事例データの条件部に、条件Cの否定条件が全く含まれないことである。

【0049】事例データにおいて、条件Cのもとで分類結果Rが成立するとは、該事例データにおいて条件Cが成立し、分類結果がRとなっていることである。

【0050】母集団が決まれば、その上で条件Cのもとでの分類結果Rの条件付確率 $p(R|C)$ が式(2)のように定義される。

【0051】

【数3】

$$p(R|C) = \frac{C \text{のもとで} R \text{を成立させる全事例データの観測確率}}{C \text{を成立させる全事例データの観測確率}} \quad \dots (2)$$

【0052】これは、分類結果の条件つき確率とも呼ばれる。

【0053】例えば、表5の母集団のもとでは、 $p(\text{魚} | \text{水中生活}) = 1.0 \quad \dots (3)$ である。

【0054】母集団に存在する全ての条件の組合せのもとでの分類結果の条件つき確率を全て正しく求める事ができれば、分類ルールの生成処理は成功裡に完了するが、これは実際上簡単なことではない。まず、条件付確率の値を正確に求めるためには多量の事例データが必要である。つぎに、条件要素がn個ある場合、条件の組合せの総数は $2^n - 1$ 個となり極めて多量の条件付確率を処理しなければならないからである。

【0055】そこで、本発明では、統計的に信頼性の高い条件付確率を優先的に用いて、分類ルールを生成している。

【0056】本発明では、事例データは1件ずつ投入され、そのたびに事例データベースの更新と分類ルールの生成・修正が行われる。この分類ルールの生成と修正を行う処理を「分類ルール生成処理」と呼ぶ事にする。分類ルール生成処理においては、投入された1件の該事例データが中心的役割をはたす事になる。そこで、この該事例データを特に「帰納事例」と呼ぶ。

【0057】分類ルール生成処理の目的は、与えられた事例データ集合の条件と分類結果の確率的関係を分類ルールとして形式化することである。このため、該帰納事例に関して、条件部を分類手がかりとして用いて、分類結果の確率値を高く導き、帰納事例の分類結果に相反する分類結果の確率値を低く導くような分類ルールを生成することが分類ルール生成処理の目標となる。

【0058】<事例データの頻度>: 入力された帰納事例は事例データベースに蓄積されてゆく。既に入力された事例データに対応して事例データの頻度が決まる。この頻度は、一般に0以上の実数値を取る。事例データベース上には概念的には以下のような表が作成されている。ただし、実体は表ではなく後述するネットワークである。

【0059】

【表7】

| 事例名称 | 頻度 | 分類結果 | 条件部 |
|------|----|------|-------------|
| 鱈3 | 1 | 魚 | 水中生活 A えら呼吸 |
| 鮭1 | 2 | 魚 | 水中生活 A 潮上産卵 |
| 鰐1 | 1 | 獣 | 陸上生活 A 肺呼吸 |
| 鰻2 | 2 | 魚 | 水中生活 A 小型 |

【0060】<分類ルール>: 分類ルールは条件部C、分類結果R、条件部Cのもとで分類結果Rが成立する条

件つき確率 $p(R|C)$ の推定値、および条件部頻度からなる。

【0061】分類ルールの例：

| ルール番号 | 分類結果 | 条件部 | 確率値 | 条件部頻度 |
|-------|------|------|------|-------|
| R1 | 魚 ← | 水中生活 | 0.86 | 5 |

【0063】分類ルールの2つの数値、条件つき確率の推定値（例中の確率値）と条件部頻度を説明するために頻度に関する2つの用語（条件部頻度、ルール成立頻度）を定義する。

【0064】<条件部頻度>：ある分類ルール r の条件部を C 、事例データの条件部を D とするとき、 D のもとで C が成り立つような全ての事例データの頻度の総和を r の条件部頻度と呼ぶ。

【0065】例：

事例データベース：

【0066】

【表9】

| 事例名称 | 頻度 | 分類結果 | 条件部 |
|------|----|------|-------------|
| 鰐3 | 1 | 魚 | 水中生活 ∧ えら呼吸 |
| 蛙1 | 2 | 魚 | 水中生活 ∧ 洲上産卵 |
| 鰐1 | 1 | 獣 | 陸上生活 ∧ 肺呼吸 |
| 鰐2 | 2 | 魚 | 水中生活 ∧ 小型 |

【0067】分類ルール：

$$p(R|C) = \frac{R \text{ かつ } C \text{ の成立する事例数 (ルール成立頻度) } + 1}{C \text{ の条件部頻度 } + n} \quad \dots (5)$$

【0073】 $t=1$, $n=2$ かつ $t=0.5$, $n=1$ などのパリエーションがあるが、特に所らないかぎり、 $t=1$, $n=2$ を用いる。分類ルールの条件つき確率は、式(2)により定義した母集団により決まる理想的条件付確率を事例データ集合より推定するものである。

【0074】事例データを用いて、確率的な分類ルールを生成し、自動分類の過程で確率値の推定を行うことにより、分類結果に確実さの定量化として確率値を付与することができる。

【0075】分類ルールの生成時に否定条件を含む分類ルールを生成する事が分類精度の向上に有効ことがある。例えば、次のような事例集合が与えられたとしよう（これまで例として登場した事例集合は無視して、この事例集合のみが与えられたとする）。

【0076】

【表11】

| ルール番号 | 分類結果 | 条件部 | 確率値 | 条件部頻度 |
|-------|------|------------|------|-------|
| R1 | 獣 ← | 水中生活 | 0.57 | 5 |
| R2 | 獣 ← | 水中生活 ∧ 肺呼吸 | 0.8 | 3 |
| R3 | 魚 ← | 水中生活 | 0.43 | 5 |
| R4 | 魚 ← | 水中生活 ∧ 肺呼吸 | 0.2 | 3 |

【0082】となる（確率値はラプラスの推定式により計算）。この分類ルールを用いて、分類問題として条件

分類ルール：

【0062】

【表8】

| ルール番号 | 分類結果 | 条件部 |
|-------|------|------|
| R2 | 獣 ← | 水中生活 |

【0069】の条件部頻度は5である。なぜならば、水中生活をする事例が5例あるからである。条件部頻度は分類結果（上記例では「獣」）には無関係である。

【0070】<ルール成立頻度>：ある分類ルール r の条件部を C 分類結果を R 、事例データの条件部を D 、分類結果を S とするとき、 D のもとで C が成り立ち、 $R=S$... (4)

であるような全ての事例データの頻度の総和を r のルール成立頻度と呼ぶ。

【0071】本発明では、条件つき確率の推定値として式(5)のラプラスの推定式を用いている。

【0072】

【数4】

| 事例 | 分類結果 | 条件部 |
|--------|------|------------|
| ザトウ鯨 | 獣 | 水中生活 ∧ 肺呼吸 |
| イルカ | 獣 | 水中生活 ∧ 肺呼吸 |
| マッコウ鯨 | 獣 | 水中生活 ∧ 肺呼吸 |
| 鯨(ます) | 魚 | 水中生活 |
| 鯨(いわし) | 魚 | 水中生活 |

【0077】但し、

【0078】

【数5】

【0079】は論理積を表す \wedge 記号である。

【0080】以上の事例から生成される分類ルールのうち、否定条件を含まない分類ルールのみを記述すると、

【0081】

【表12】

＝「水中生活」を解くと、分類結果は、1位「獣」確率値0.57、2位「魚」確率値0.43となる。これ

は、誤っており、1位が“魚”、2位が“獣”となるのが正しい。

【0083】ここで、否定条件付の分類ルールも記述す

る。

【0084】

【表13】

| ルール番号 | 分類結果 | 条件部 | 確率値 | 条件部頻度 |
|-------|------|--------------|------|-------|
| R5 | 獣 ← | 水中生活 A → 肺呼吸 | 0.33 | 3 |
| R6 | 魚 ← | 水中生活 A → 肺呼吸 | 0.67 | 3 |

【0085】

【数6】

【0086】は否定を表す記号であり、

【0087】

【数7】

”→肺呼吸”

【0088】は、“肺呼吸”という条件が無いことを示している。

【0089】この分類ルールを用いて、分類問題として条件=“水中生活”を解くと、1位“魚”確率値0.67、2位“獣”確率値0.33と正しい順序の分類結果が得られる。

【0090】以上のように否定条件付きの分類ルールは分類精度の向上に有効であるが、その生成は否定条件なしの分類ルールほど容易ではない。それは、否定条件なしの分類ルールは帰納事例の条件を組み合わせる事により生成可能であるのに対して、否定条件付きの分類ルールでは、否定条件が帰納事例の条件に含まれておらず過去の事例集合から探索する必要があるためである。以降では、まず本発明の全体像を示し、次に否定条件なしの分類ルールの生成方法を説明し、その後に、否定条件付きの分類ルールの生成方法を示す。

【0091】

【発明の実施の形態】次に、本発明の実施の形態について図面を参照して詳細に説明する。

【0092】図1及び図2は、本発明の実施の形態に係る分類装置の構成を示すブロック図である。図1は、請求項1の発明の実施の形態に対応し、図2は、請求項3に記載の実施の形態に対応している。以下では、図2を参照して、分類ルール生成時と分類時の処理手順を説明する。図1との処理手順の違いは後述する。

【0093】図1を参照すると、本発明の実施の形態は、既知事例データおよび未知事例データを入力する入力装置103と、確率的な情報を含む分類ルールを蓄える分類ルールデータベース106と、前記既知事例データを条件部の論理関係に基づいてネットワーク化して蓄積する事例データベース105と、入力された前記既知事例データおよび前記未知事例データの条件部と前記分類ルールを用いて分類結果の確率値を推定する確率値推定装置102と、分類ルールの有効性を統計的検定により評価することにより無駄な分類ルールの生成を抑制す

る分類ルール生成装置101と、出力装置104と、を含んで構成されている。

【0094】また、図2を参照すると、図1の構成に、更に、否定条件探索装置107を備えた構成とされ、否定条件探索装置107は、入力された前記既知事例データの条件部全部あるいは条件の一部を入力として受取り、前記事例データベースを探索して否定条件を含む事例ノードを推定し、分類ルール生成装置101は否定条件探索装置107を用いて、否定条件を含む分類ルールを生成する機能を具備している。

【0095】<分類ルール生成時の処理手順>:

ステップ1: 入力装置から分類結果が既知の事例データ1個の入力を受け、分類ルール生成装置101と確率値推定装置102へ転送する。この事例データは、以降の処理で他の事例とは異なる役割を果たすため、これを特に「帰納事例」と呼ぶ。

【0096】ステップ2: 分類ルール生成装置101は、事例データベース105へ帰納事例の追加を行う。

【0097】ステップ3: 確率値推定装置102は、帰納事例の条件と分類ルールデータベース106を用いて分類結果とその確率値を求める。

【0098】ステップ4: 分類結果が外れていた場合には、分類ルールの生成を行うためステップ5へ、分類結果が正しい場合にはステップ9へ分岐する。

【0099】ステップ5: 分類ルール生成装置101は、まず、否定条件を含まない分類ルールの生成の試みを行い、生成できればステップ9へ分岐する。

【0100】ステップ6: 分類ルール生成装置101は、否定条件探索装置107へ探索条件を送る。

【0101】ステップ7: 否定条件探索装置107は、事例データベース105を探索して、検索条件に適合する否定条件を求め、分類ルール生成装置101へ転送する。

【0102】ステップ8: 分類ルール生成装置101は、否定条件付の分類ルールを生成して分類ルールデータベース106へ追加する。

【0103】ステップ9: 次の事例データがあればステップ1へ戻り、無ければ終了する。

【0104】図1に示す分類装置では、否定条件探索装置107が設けられていないため、上記ステップ6、ステップ7、ステップ8の処理が省略される。

【0105】<自動分類時の処理手順>:

ステップ1: 入力装置103から分類結果が未知の事例

データの入力を受け取り、確率値推定装置102へ転送する。

【0106】ステップ2：確率値推定装置102は該事例データの条件と分類ルールデータベースを用いて分類結果とその確率値を求める。

【0107】ステップ3：次の事例データがあればステップ1へ、無ければ終了する。

【0108】図1の分類装置でも同様である。

【0109】＜否定条件なしの分類ルール生成方法＞：否定条件なしの分類ルールは、原則として事例の条件を組み合わせる事により生成する事ができる。

| ルール番号 | 分類結果 | 条件部 | 確率値、条件部頻度は省略 |
|-------|------|-----------------|--------------|
| R1 | 獣 ← | 水中生活 | |
| R7 | 獣 ← | 肺呼吸 | |
| R9 | 獣 ← | 巨大 | |
| R2 | 獣 ← | 水中生活 ∧ 肺呼吸 | |
| R8 | 獣 ← | 肺呼吸 ∧ 巨大 | |
| R10 | 獣 ← | 巨大 ∧ 水中生活 | |
| R11 | 獣 ← | 水中生活 ∧ 肺呼吸 ∧ 巨大 | |

【0114】ただし、すべての分類ルールを生成すると総ルール数が爆発的に増大する為、現実的な応用に際しては、有益な分類ルールのみを生成するようにコントロールする事が重要である。生成可能な分類ルール数は、事例の条件数を n とすると、その組合せの総数 $2^n - 1$ だけ存在する。例えば、 $n = 10$ のとき、生成可能な分類ルール数は1023となる。 n が大きくなるにつれてこの数は爆発的に増大するため、有益な分類ルールのみを生成するような手段が必要になる。

【0115】本発明においては、有益な分類ルールのみを生成するため以下の2つの手段を用いている。

【0116】1. 第1の手段として、統計的検定によって分類ルールの有効性を判定する。これは請求項1に対応している。

【0117】2. 第2の手段として、帰納事例がうまく分類できた場合は、新しい分類ルールは生成しない。これは請求項2に対応している。

【0118】以下では、統計的検定による分類ルールの有効性判定手順を説明する。

【0119】＜分類ルールの有効性＞：分類ルールの有効性は観念的には以下のように定義される。

【0120】1. 分類ルール r の分類結果を R 条件部を C とするとき、条件つき確率の $p \sim (R | C)$ が、母集団から決まる条件つき確率 $p(R | C)$ と近い値を取り、

2. 分類ルール r を分類ルールベースに追加する前の分類システムの分類能力と追加した後の分類能力が大きく異なるとき、分類ルール r を有効とする。

【0121】第1の条件は、条件つき確率 $p(R | C)$ の推定にラプラスの推定式(5)を用いる事によりある程度満足されている。第2の条件の判定方法を以下に述

【0110】例：

【0111】

【表14】

| 事例 | 分類結果 | 条件部 |
|------|------|-----------------|
| ザトウ鯨 | 獣 | 水中生活 ∧ 肺呼吸 ∧ 巨大 |

【0112】に対しては、以下の分類ルールが生成可能である。

【0113】

【表15】

べる。この判定では、既存の分類ルールベースから追加しようとする分類ルールが導けるかどうかが焦点となる。

【0122】＜ルール有効性判定手順＞：

ステップ1：帰納事例を事例データベースに追加する。

【0123】ステップ2：該事例データベースを用いて、有効性判定対象の分類ルールの条件部頻度とルール成立頻度を得る。以下では、有効性判定対象の分類ルールを「当該分類ルール」と呼ぶ。

【0124】ステップ3：既存の分類ルールベースを用いて、当該分類ルールの条件部 C を分類手がかりとして自動分類する（自動分類の1例として、特開平7-200545号公報に記載の確率的演繹推論装置による分類が用いられる）。その結果、当該分類ルールの分類結果 R の条件つき確率の推定値 $p \sim (R | C)$ が得られる。

【0125】ステップ4：該条件つき確率推定値 $p \sim (R | C)$ のもとで該条件部頻度と該ルール成立頻度以上の異常な状態が起こる確率 q を計算する。

【0126】ステップ5：該確率 q があらかじめ設定された危険率以上である場合、当該分類ルールは既存の分類ルールベースから自然に導けることがわかる。そこで、当該分類ルールは有効でない判定する。

【0127】逆に、 q があらかじめ設定された危険率未満である場合、当該分類ルールは既存の分類ルールベースから自然に導けるものではなく特殊なものであることがわかる。そこで、当該分類ルールを有効と判定する。

【0128】例：有効性判定対象の分類ルールを、

【0129】

【表16】

| ルール番号 | 分類結果 | 条件部 |
|-------|------|------------|
| R2 | 獣 ← | 水中生活 ∧ 肺呼吸 |

【0130】とする。

【0131】まず、既存の分類ルールベースを用いて、

条件付確率

【0132】

【数8】

$p(\text{獣} | \text{水中生活} \wedge \text{肺呼吸})$

| ルール番号 | 分類結果 | 条件部 | 確率値 | 条件部頻度 |
|-------|------|------|-------|-------|
| R1 | 獣 ← | 水中生活 | 0.019 | 105 |
| R7 | 獣 ← | 肺呼吸 | 0.90 | 68 |

【0138】これらの分類ルールの確率値を用いて、条件付確率

【0139】

【数10】

$p(\text{獣} | \text{水中生活} \wedge \text{肺呼吸})$

【0140】を推定する。これには幾つかの推定方法があるが、1例として、特開平7-200545号公報に記載の確率的演繹推論装置をあげる。

【0141】詳細は省略するが、既存の分類ルールベースと推定手段を用いて分類手がかかり

【0142】

【数11】

「水中生活 ∧ 肺呼吸」

【0143】のもとで分類結果「獣」が出現する確率の推定値

【0144】

【数12】

$p(\text{条件付確率 } p(\text{獣} | \text{水中生活} \wedge \text{肺呼吸}))$

【0145】が0.2と推定されたとする。

【0146】また、事例データベースから得られた当該分類ルールの頻度は、

| ルール番号 | 分類結果 | 条件部 | 確率値 | 条件部頻度 |
|-------|------|------------|------|-------|
| R2 | 獣 ← | 水中生活 ∧ 肺呼吸 | 0.83 | 10 |

【0152】一方、確率値 q が危険率を上回る値であった場合は、有効とは判断できないため、分類ルールには追加されない。なお、危険率は統計的検定に用いられる閾値であり、帰無仮説を誤って棄却する確率である。分類ルールの有効性判定においては、無効な分類ルールを

【0133】を推定する。そのため、

【0134】

【数9】

条件部 (水中生活 ∧ 肺呼吸)

【0135】のもとで成立する分類ルールの集合を求める。

【0136】例：

【0137】

【表17】

条件部頻度 = 10,

ルール成立頻度 = 9

であるとする。

【0147】このとき、生成しようとする分類ルールの成立確率を p と仮定した基で、事例データベースから求めた頻度以上に異常な状態が起こる確率 q は、

【0148】

【数13】

$$\begin{aligned}
 q &= {}_{10}C_9 * (1-p) * p^9 + p^{10} \\
 &= 10 * 0.8 * 0.2^9 + 0.2^{10} \\
 &= 0.0000042
 \end{aligned}$$

【0149】と極めて低い。

【0150】この確率値があらかじめ設定される危険率 (0.05, 0.1等) を下回った場合、生成しようとする分類ルールは既存の分類ルールとは異ると判定される。これは生成しようとする分類ルールが有効であることを示しているので、ラプラスの推定式 (5) を用いて確率値を求め、分類ルールベースに追加する。追加される分類ルールを以下に示す。

【0151】

【表18】

生成してしまう確率である。

【0153】以上では概念的な説明のために確率値 q を計算して危険率と比較する統計的検定方法を示した。しかし統計的検定には様々な方法があり、本発明では発明の請求範囲を上記の確率計算法に限らず、適用可能な全

ての統計的検定方式を用いる。

【0154】＜否定条件つき分類ルールの生成手順＞：前述のように適切な否定条件つき分類ルールは分類精度を向上させる。以下ではその生成手順を説明するが事例データベースの探索を無駄無く行うことにより計算時間の短縮と記憶装置の負担を軽減している。これは請求項3に対応している。

【0155】否定条件つき分類ルールの生成手順では、ステップ1：追加すべき否定条件を事例データベースから探索し、分類ルールを生成し
ステップ2：統計的検定を行い分類ルールの有効性の判定を行い、
ステップ3：有効と判定されたものを分類ルールベースに追加する。

【0156】否定条件の探索では、事例データベースを条件の簡単なものから複雑なものへとたどりながら、否定条件となり得る条件を探索するが、事例データベースは条件の論理的関係に基いたネットワークを構成しており、これが探索時間の短縮を実現している。統計的検定による分類ルールの有効性判定は否定条件なしの場合と同様である。

【0157】否定条件の探索を説明する前に、事例データベースの構成を説明する。事例データベースの構成要素を、事例ノードと呼び、図3に示すように、一意番号、条件、検査マーク、下位事例リスト、事例情報から構成する。以下にその意味を示す。

【0158】一意番号：事例ノードを一意に示す番号。

【0159】条件：事例ノードの条件部を記録する。

【0160】検査マーク：否定条件の探索などの処理において、この事例ノードが既に処理されたかどうかを記録する。

【0161】下位事例リスト：下位事例ノードを記録

したリスト。“下位事例ノード”の意味は後述する。

【0162】事例情報：事例ノードの条件のもとで、どのような分類結果が何回現れたかを記録する。

【0163】＜下位の事例ノード＞：事例ノードEの条件部をCとすると、条件部Cの条件を全て含む事例ノードの集合をSとする。Sの中で、他の事例の条件を完全に含むことはない（もともと簡単な条件部を持つ）事例ノードをEの“下位事例ノード”あるいは“下位ノード”と呼ぶ。下位事例ノードは複数あり得る。

【0164】例：事例ノードとその条件部が以下の表の時、E0の下位事例ノードはE1、E2である。E4はE0の下位事例ノードにはならない。

【0165】

【表19】

| 事例番号 | 条件部 |
|------|------------------|
| E0 | 水中生活 |
| E1 | 水中生活 ∧ えら呼吸 |
| E2 | 水中生活 ∧ ひれ |
| E3 | えら呼吸 |
| E4 | 水中生活 ∧ えら呼吸 ∧ ひれ |

【0166】事例情報の例：事例の集合、

【0167】

【表20】

| 事例 | 分類結果 | 条件部 |
|--------|------|------|
| ザトウ鯨 | 鯨 | 水中生活 |
| イルカ | 鯨 | 水中生活 |
| マッコウ鯨 | 鯨 | 水中生活 |
| 鯨(ます) | 魚 | 水中生活 |
| 鯨(いわし) | 魚 | 水中生活 |

【0168】は、事例ノード

【0169】

【表21】

| 一意番号 | 条件部 | 検査マーク | 下位事例リスト | 事例情報 |
|------|------|-------|---------|------------|
| E1 | 水中生活 | 未検査 | なし | 鯨(3), 魚(2) |

【0170】に対応する。この事例情報は、条件“水中生活”に対して分類結果“鯨”の事例が3件、“魚”の事例が2件存在した事を記録している。

【0171】同位事例ノード：ある事例ノードは複数の下位事例ノードを持つことがあるが、これらの下位事例ノードを互いに“同位事例ノード”あるいは“同位ノード”と呼ぶ。

【0172】上位事例ノード：ある事例ノードEを下位事例ノードとする事例ノードFはEから見て“上位事例ノード”あるいは“上位ノード”と呼ばれる。

【0173】事例データベース105は、事例の探索の速度を効率化するため、個別条件毎に、探索エントリポイントという事例ノードのリストを生成する。

【0174】探索エントリポイント：探索エントリポイントは1つの条件に対して1個存在しており、その条件

を含む最も簡単な事例ノードのリストで構成される。

【0175】事例が多数与えられると、事例ノードは連結され、ネットワークを形成する。以下の事例が与えられた後の事例ノードの連結例を図4に示す。

【0176】

【表22】

| 事例 | 分類結果 | 条件部 |
|--------|------|------------|
| ザトウ鯨 | 鯨 | 水中生活 ∧ 肺呼吸 |
| イルカ | 鯨 | 水中生活 ∧ 肺呼吸 |
| 鯨(ます) | 魚 | 水中生活 |
| 鯨(いわし) | 魚 | 水中生活 |
| 肺魚 | 魚 | 水中生活 ∧ 肺呼吸 |
| 鰐 | 鯨 | 肺呼吸 |
| 牛 | 鯨 | 肺呼吸 |

【0177】事例データの入力とともに上位、下位の連

結を更新しながら事例データベースが形成される。この事例データベースを用いれば、ある条件と分類結果が与えられた際に、既知の事例中の条件の成立頻度と分類結果の成立頻度を効率的に求めることができる。

【0178】事例ノードEを含んでそれ以下の下位の事例ノードの条件成立頻度の総和をEの条件総頻度、分類結果Rの総成立頻度を条件EにおけるRの成立総頻度と呼ぶ。条件総頻度はEの条件以外は無視した際の事例の個数であり、Rの成立総頻度は同じくRの成立する事例の個数である。これらの値は、事例ノードに記憶しておいてもよい。

【0179】〈否定条件の検索手順〉：否定条件は、前述したように帰納事例（分類ルール生成の手がかりとなる1件の事例）に含まれていないので、事例データベースを検索して発見しなければならない。その具体的な手順は、図5に示されているが、ここでは概念的に説明する。

【0180】否定条件の検索では、分類結果Rと検索条件Cが当らる。これが、図2に示した分類装置のブロック図において、分類ルール生成装置101から否定条件探索装置107へ送られる情報である。

【0181】検索条件としては、

1. 帰納事例の条件
2. 帰納事例を自動分類する際に利用される既存の分類ルールの条件

【0182】2の肯定条件は1の部分集合であり、検索時間の短縮効果がある。この検索条件を用いて否定条件の検索を行う。以下の手順は、分類結果Rを高い確率で積極的に導く分類ルールを生成する場合を想定する。分類結果Rを低い確率で否定的に導く分類ルールを生成する場合は、確率値や検定の考え方が逆になる。

【0183】探索は、検索条件の一部を含む検索条件と矛盾しないような事例ノードのうち、条件の少ない事例ノードから始めて、下位の事例ノードへと進められる。ある事例ノードにおいて、以下の条件が満たされた場合、この事例ノードには否定条件が含まれると推定する。

【0184】〔否定条件を含む事例ノードの条件〕：ある事例ノードEの条件総頻度を n 、分類結果Rの成立総頻度を s とし、Eの上位事例ノードFにおける条件総頻度を m 、分類結果Rの成立総頻度を t とする。あるFについて、 t/m が十分高く、 s/n が十分低く、 t/m と s/n の間に有為な差がある場合、事例ノードEは否定条件を含むと推定する。

【0185】否定条件を含むと推定された事例ノードの条件Dと帰納事例の条件の差分を否定条件とする。

【0186】〈ソフトウェア関連発明に関する説明〉：図8は、請求項4の処理手順記録媒体を含む分類装置のブロック図である。図8を参照すると、本発明の分類装置は、図1に示した構成に加えて、分類ルール生成装置

駆動プログラムを記録した処理手順記憶媒体108を備える。この処理手順記録媒体108は磁気ディスク、半導体メモリその他の記録媒体であってよい。また、分類ルール生成装置駆動プログラムに係わる分類ルール生成装置101は、請求項4でいうコンピュータにより実現される。

【0187】分類ルール生成装置駆動プログラムには、既に述べた、分類ルール生成時の処理手順、自動分類時の処理手順、ルール有効性判定手順が含まれている。これらは、分類装置の初期設定時に、処理手順記録媒体108から分類ルール生成装置101をなすコンピュータに読み込まれ、実行される。

【0188】図9は、請求項5の処理手順記録媒体を含む分類装置のブロック図である。図9によると、本発明の分類装置は、図2に示した構成に加えて、装置駆動プログラムを記録した処理手順記憶媒体108を備える。この処理手順記録媒体108は、磁気ディスク、半導体メモリその他の記録媒体であってよい。また、この装置駆動プログラムに係わる分類ルール生成装置101と否定条件探索装置107は、請求項5でいうコンピュータにより実現される。

【0189】この装置駆動プログラムには、既に述べた、分類ルール生成時の処理手順、自動分類時の処理手順、ルール有効性判定手順、否定条件つき分類ルールの生成手段、否定条件の検索手順が含まれている。これらのうち、分類ルール生成時の処理手順、自動分類時の処理手順、ルール有効性判定手順、否定条件つき分類ルールの生成手順は、分類装置の、初期設定時に、処理手順記録媒体108から分類ルール生成装置101をなすコンピュータに読み込まれ、既に述べた手順により実行される。

【0190】また、否定条件の検索手順は、分類装置の初期設定時に、処理手順記録媒体から否定条件探索装置108をなすコンピュータに読み込まれ、既に述べた手順により実行される。

【0191】

【実施例】上記した本発明の実施の形態について更に詳細に説明すべく、本発明の実施例について以下に説明する。

【0192】図1は、請求項1の発明の一実施例に係る分類装置の構成を示すブロック図である。図2は、請求項3の発明の一実施例に係る分類装置のブロック図である。図1を参照すると、本実施例に係る分類装置は、既知事例データおよび未知事例データを入力する入力装置103と、確率的な情報を含む分類ルールを蓄える分類ルールデータベース106と、前記既知事例データを条件部の論理関係に基づいてネットワーク化して蓄積する事例データベース105と、入力された前記既知事例データおよび前記未知事例データの条件部と前記分類ルールを用いて分類結果の確率値を推定する確率値推定装置

102と、分類ルールの有効性を統計的検定により評価することにより無駄な分類ルールの生成を抑制する分類ルール生成装置101と、出力装置104と、を含んで構成されている。

【0193】また図2を参照すると、この分類装置は、図1の構成に、更に、否定条件探索装置107を備えた構成とされ、否定条件探索装置107は、入力された前記既知事例データの条件部全部あるいは条件の一部を入力として受取り、前記事例データベースを探索して否定条件を含む事例ノードを推定し、分類ルール生成装置101は否定条件探索装置107を用いて、否定条件を含む分類ルールを生成する機能を具備している。

【0194】これらの分類装置における分類ルールの生成手順を以下に示す。

【0195】＜分類ルールの生成手順＞：図5は、分類ルールの生成手順の一実施例を示す流れ図である。装置名称については、図2に示した分類装置のブロック図に対応している。本実施例は、請求項1、2、3全てに対応しているが、各請求項に限定する場合は後述する。

【0196】ステップ（Step）1：入力装置103から事例データ（帰納事例）1個の入力を受け、分類ルール生成装置101と確率値推定装置102へ転送する。

【0197】ステップ2：分類ルール生成装置101は事例データベース105へ帰納事例の追加を行う。

【0198】ステップ3：確率値推定装置102を用いて帰納事例の条件部と分類ルールから分類結果とその確率値を求める。

【0199】ステップ4：各々の分類結果について、推定した確率値と正解確率値が設定された閾値以上に隔たっている場合には、分類結果が外れているものと判定する。推定が外れた分類結果は全て“推定結果誤りリスト”に記録する。

【0200】ステップ5：推定結果誤りリストが空であれば、ステップ1へ移り、そうでなければ、ステップ6の処理へ進む。

【0201】ステップ6：推定結果誤りリストの先頭の分類結果が、ステップ6からステップ12までの処理の対象となる分類ルールの分類結果となる。確率値推定装置102から、ステップ3の分類時に用いた分類ルールを、分類ルール生成装置101に転送する。

【0202】ステップ7：帰納事例の条件を各個の条件単位に分解し追加条件集合Sを生成する。

【0203】

【数14】

例：帰納事例の条件 = 水中生活 A 肺呼吸

【0204】追加条件集合S = 水中生活、肺呼吸

【0205】ステップ8：該分類ルール各々に該追加条件集合Sの要素である各個の条件を追加して、新しい分

類ルールの条件を生成する。追加すべき条件の個数の上限値は、装置起動時に外部から与えられる。

【0206】ステップ9：ステップ8で生成した新分類ルールの条件部頻度、ルール成立頻度を事例データベースより求め、ステップ3で求めた確率値pを用いて、前記本発明の発明の実施の形態で説明し、た統計的検定により該新分類ルールの有効性を判定する。該新分類ルールが有効であればステップ11へ、無効ならばステップ10へ進む。

【0207】ステップ10：生成の対象となる新しい分類ルールが残っていればステップ8へ、さもなければ、ステップ12へ進む。

【0208】ステップ11：該新分類ルールを分類ルールデータベース106に追加する。

【0209】ステップ12：否定条件を含む分類ルールの生成を試みる。

【0210】ステップ13：分類結果誤りリストの先頭分類結果を除去する。除去された分類結果はステップ6からステップ13までの処理で分類ルールの分類結果となって来たものである。分類結果誤りリストの要素が残っていればステップ6へ。

【0211】ステップ14：事例データが残っていれば、ステップ1へ、さもなければ終了する。

【0212】なお、請求項1の実施例に限定する場合は、上記の分類ルールの生成手順に、以下に記載の修正を行う。また、対応するブロック図は、図に示す構成とされる。

【0213】修正：

ステップ4：自動分類の結果が誤りであった場合はすべての分類結果を“推定結果誤りリスト”に記録する。

【0214】ステップ12：行わない。

【0215】自動分類の結果が誤りであるかどうかの判定には以下のような方法がある。

【0216】・確率値の高い順に自動分類の分類結果を並べて、その1位が帰納事例の分類結果と一致していなければ誤りと判断する。

【0217】・確率値の高い順に自動分類の分類結果を並べて、その1位の確率値があらかじめ設定した閾値を超え、2位の確率値が別にあらかじめ設定した閾値を下回る場合、1位が帰納事例の分類結果と一致していなければ誤りと判断する。

【0218】・自動分類の結果出力された上位の分類結果の確率値を総和が1になるように正規化し、1位の確率値があらかじめ設定した閾値を超えた場合、1位が帰納事例の分類結果と一致していなければ誤りと判断する。

【0219】請求項1、2に限定する場合は、上記の分類ルールの生成手順に以下に記載の修正を行う。また、対応するブロック図は、図1となる。

【0220】修正：

ステップ12:行わない。

【0221】<否定条件を含む分類ルールの生成手順>:これは、請求項3の分類ルール生成装置における分類ルール生成手順である。

【0222】図6は、固定条件を含む分類ルールの生成手順の一実施例である。これは5のステップ12で利用される手順の一実施例の流れ図である。

【0223】ステップ1:自動分類に用いられた分類ルールのうちから望ましくない確率値を持つ分類ルールの条件をすべて集める条件集合Sを生成する。

【0224】望ましくない確率値を持つ分類ルールとは、

1. 正しい分類結果を導くべき分類ルールの場合、確率値が設定された閾値よりも有意に低いものであり、
2. 誤った分類結果を導くべき分類ルールの場合、確率値が設定された閾値よりも有意に高いものである。

【0225】ステップ2:該条件集合Sを論理的に成立させるような事例ノードを全て集め、これを事例集合Eとする。

【0226】ステップ3:Eの要素である事例ノード各々について、否定条件を含んでいるかどうかの推定をおこない、得られた否定条件を否定条件集合Nに累積する。

【0227】この推定方法は、上記発明の実施の形態で説明されている。

【0228】ステップ4:該条件集合Sに該否定条件集合Nを追加して新しい分類ルール候補hの条件部とする。

【0229】ステップ5:該分類ルール候補hの有効性を統計的検定により判定する。その際、事例データベースからhの条件成立頻度とルール成立頻度を求める。

【0230】ステップ6:有効であれば該分類ルール候補hを分類ルールとして生成する。

【0231】図7は、否定条件の探索手順の一実施例の流れ図である。これは、「否定条件を含む分類ルールの生成手順」および図6のステップ2で用いられた事例ノード探索手順でもある。

【0232】図7において、開始時点では、検索条件と分類結果が与えられる。また、図7中の用語の意味は以下の通りである。

【0233】ノードスタック:検索対象の事例ノードを一時格納するスタック。単に「スタック」と呼ぶ場合もある。一般にスタックとは、先入れ、後出し式の記憶装置である。

【0234】復帰状態スタック:スタックと連動するスタックで、ノードスタックの先頭ノードを処理し終えて次の事例ノードを処理する際にどのステップから始めるかを示す。

【0235】回答ノードリスト:探索手順が回答として返す事例ノードのリスト。

【0236】ステップ1:初期設定を行う。検索条件を個別の条件単位に分解し、各条件のエントリーポイントの事例ノードを順にノードスタックに積む。最初の事例ノードを積んだ際に復帰状態スタックに「終了」を積む。つぎの事例ノードからは1個積むと同時に復帰状態スタックに「ステップ2」を1個積む。

【0237】ステップ2:探索開始。ノードスタックの先頭事例ノードを探索対象ノードとする。

【0238】ステップ3:探索対象ノードは「未検査」状態か判定し、未検査ならばステップ4へ、検査済ならばステップ12へ進む。

【0239】ステップ4:探索対象ノードを「検査済」とする。

【0240】ステップ5:探索対象ノードの基で検索条件は成立するか判定し、成立するならばステップ6へ、さもなければステップ8へ進む。

【0241】ステップ6:探索対象ノードは否定条件となりうるような条件を含んでいるか判定し、含んでいればステップ7へ、含んでいなければステップ8へ進む。

【0242】ステップ7:探索対象ノードを回答ノードノストに追加する。

【0243】ステップ8:探索対象ノードに同位ノードは存在するか判定し、存在するならばステップ9へ、存在しない場合ステップ10へ進む。

【0244】ステップ9:同位ノードをスタック先頭に積む。さらに復帰状態スタックに「ステップ10」を積む。ステップ3へ跳ぶ。

【0245】ステップ10:探索対象ノードの条件総頻度が設定された閾値を下回っていた場合、下位ノードの探索は行わない。下位ノードの探索を行う場合ステップ11へ、行わない場合ステップ12へ進む。

【0246】ステップ11:下位ノードをノードスタック先頭に積む。復帰状態スタックに「ステップ12」を積み、ステップ3に跳ぶ。

【0247】ステップ12:スタック先頭ノードを除去する。その結果2番目の事例ノードがあれば、これがスタック先頭ノードとなり、同時に探索対象ノードとする。

【0248】ステップ13:復帰状態スタックの先頭要素を復帰変数に退避し、先頭要素を除去する。

【0249】ステップ14:復帰変数=「ステップ10」ならばステップ10に跳ぶ。

【0250】ステップ15:復帰変数=「ステップ12」ならばステップ12に跳ぶ。

【0251】ステップ16:復帰変数=「終了」ならば終了に跳ぶ。

【0252】

【発明の効果】以上に示したように、本発明によれば、現実的な分類問題に対して有効な自動分類を行うために、分類ルールの生成時にそのルールの有効性を判定

し、無駄な分類ルール生成を防ぐ手段を備え、条件の論理的関係に基づいてネットワーク化された事例データベースを探索することにより、否定条件を発見し、否定条件付きの分類ルールを生成する手段を備え、第1の手段により、分類ルールの記憶のために必要な記憶装置の総量を減少させることができ、分類ルールの理由・操作の処理時間を短縮させることができる、という効果を奏する。これは、現実問題を解く際に有効である。

【0253】また、本発明によれば、第2の手段により、全事例を探索するよりも短い処理時間で否定条件の探索を行うことができ、否定条件付きの分類ルールを効率的に生成することができるという効果を奏する。否定条件付きの分類ルールは自動分類の精度向上につながる。

【図面の簡単な説明】

【図1】本発明の一実施例の構成を示すブロック図である。

【図2】本発明の別の実施例の構成を示すブロック図である。

【図3】本発明の実施例を説明するための図であり、事例ノードの概念図である。

| 一意番号 | 条件 | 検査マーク | 下位事例リスト | 事例情報 |
|------|----|-------|---------|------|
|------|----|-------|---------|------|

【図4】本発明の実施例を説明するための図であり、連結した事例ノードを示す例を表す図である。

【図5】本発明における分類ルールの生成の手順の一実施例を示す図である。

【図6】本発明において否定条件を含む分類ルールの生成の手順の一実施例を示す図である。

【図7】本発明において否定条件を含む事例ノードの探索手順の一実施例を示す図である。

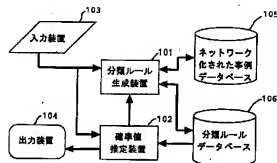
【図8】本発明の処理手順記録媒体を含む分類装置のブロック図を示す図である。

【図9】本発明の処理手順記録媒体を含む分類装置ブロック図を示す図である。

【符号の説明】

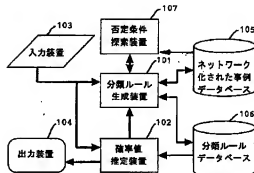
- 101 分類ルール生成装置
- 102 確率値推定装置
- 103 入力装置
- 104 出力装置
- 105 事例データベース
- 106 分類ルールデータベース
- 107 否定条件探索装置
- 108 処理手順記録媒体

【図1】

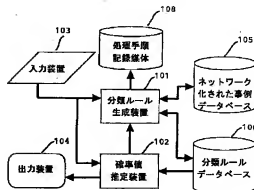


【図3】

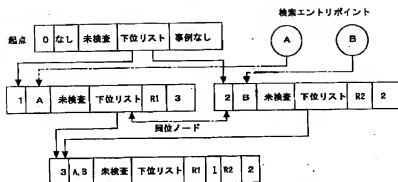
【図2】



【図8】

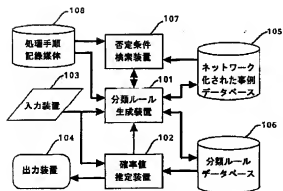


【図4】

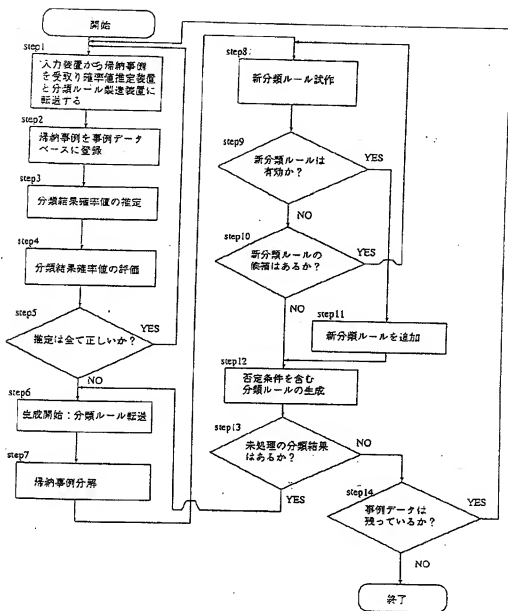


上例は、事例として、(結論、条件)のペアが、(R1, A)が3件、(R2, B)が2件、
 (R1, A, B)が1件、(R2, A, B)が2件与えられたことを示す
 A=水中生活 B=肺呼吸 R1=魚 R2=獣である

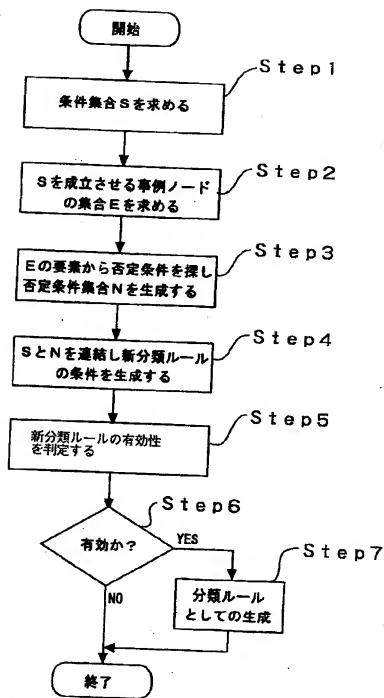
【図9】



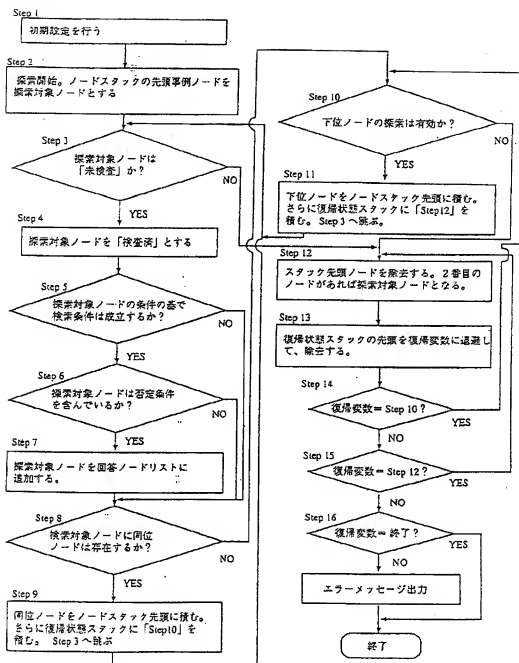
【図5】



【図6】



【図 7】



- (56) 参考文献 特開 平 6-44206 (J P, A)
特開 平 5-35484 (J P, A)
特開 平 7-200545 (J P, A)
斉藤、中野、「ノイズを含む事例からのルール抽出：R F 3 アルゴリズム」、情報処理学会論文誌、V o l. 33、N o. 5、社団法人情報処理学会・発行 (1992年)、p p. 636~644 (特許庁 C S D B 文献番号：C S N T 199800280007)
吉田、山村、小林、「属性の識別能力の局所性を考慮した確率的決定木の構築」、人工知能学会誌、V o l. 11、N o. 2、社団法人人工知能学会・発行 (1996年 3 月)、p p. 264~272 (特許庁 C S D B 文献番号：C S N T 199800390010)
谷澤、上原、前川、「典型性に基づく概念学習アルゴリズム」、情報処理学会研究報告、V o l. 93、N o. 5 (93-A I-86)、社団法人情報処理学会・発行 (1993年)、p p. 33~40 (特許庁 C S D B 文献番号：C S N T 199800022005)

- (58) 調査した分野(Int. Cl. 7, D B 名)
G06F 9/44
G06F 17/30
J I C S T ファイル (J O I S)
C S D B (日本国特許庁)